Modeling Connected Customer Lifetime Value Modeling Connected Customer Lifetime Value (CCLV) in the Banking Domain Full Paper

Conference Paper · August 2017

Modeling Connected Customer Lifetime Value (CCLV) in the Banking Domain

Full Paper

Mark Lycett
Royal Holloway, University of London
Mark.Lycett@rhul.ac.uk

Alaa Marshan
Brunel University London
Alaa.Marshan@brunel.ac.uk

Abstract

Customer Lifetime Value (CLV) has become increasingly important as a marketing metric because of its ability to predict the future profitability of clients, potentially enabling more appropriate marketing strategies. Traditional CLV models, however, do not reflect the (dynamic) networks of business transactions. This research develops a Connected Customer Lifetime Value (CCLV) model based on an empirical analysis of transactions in the financial services domain. The model was applied to a significant number of transactions between firms using a modern open source computing infrastructure (Spark plus Hadoop). We have illustrated the outcomes of the application of the model via a 'top and bottom 20' listing of firms in relation to their value network. In positive terms, application of the model allows our research partner to see the network implications of decisions they make with respect to customers and opens up an arena for innovation re network-based products and services.

Keywords (Required)

Business Analytics, Big Data, Connected Customer Lifetime Value (CCLV), Social Network Analysis.

Introduction

The consequences of the world financial crisis in 2008 have impacted negatively on the margins and profitability of financial service organizations. In part as a response, these organizations have worked to innovate and improve techniques that differentiate and value their clients, so that they can better market their banking products and target the (dynamic) needs of their clients. The need for new measures that value customer performance stems from the fact that traditional approaches for marketing and targeting customers are neither powerful nor precise enough to measure and predict the return on marketing investment. Additionally, aggregated financial metrics (such as aggregated profitability) have limited diagnostic capability and provide only a generalized understanding on customer profitability (Gupta et al. 2006). Consequently, disaggregate methods have received increasing attention.

Customer Lifetime Value (CLV) is one such measure that has become increasingly important (Gupta et al. 2006; Mzoughia and Limam 2015). CLV is a disaggregate metric that can be used to identify profitable customers and allocate resources accordingly, providing more customer's specific insights (Gupta et al. 2006; Kumar et al. 2006). Prior studies have argued CLV to be a key concept for any business that can positively affect its current and future performance (Baum and Singh 2008; Feiz et al. 2016; Kumar et al. 2006). Unsurprisingly, therefore, CLV is a now fundamental concept in many customer relationship management approaches, such as one-to-one, loyalty, and database marketing (Blattberg et al. 2009; Borle et al. 2008). One of the main strengths of CLV analysis is that it can be used to predict the future profitability of clients, leading to more accurate marketing strategies and decisions relating to customers (Chang et al. 2012). In summary, the usefulness of CLV models for selecting and targeting specific customers is justified by "customers who are selected on the basis of their lifetime value provide higher profits in future periods than do customers selected on the basis of several other customer-based metrics" (Venkatesan and Kumar 2004, P 106).

Crudely speaking, traditional CLV models consider the profit generated by the customer while subtracting the cost of customer acquisition and retention. Though valuable, this approach does not reflect the dynamic networks that reflect the business relationships of our time. Thus, the concept of the 'value network' is important to consider, since it potentially reflects both tangible and intangible dynamic value exchange in a network of customers transacting among each other (Hosseini and Albadvi 2010). Fledgling work has been done in this area (Hosseini and Albadvi 2010; Klier et al. 2014; Weinberg and Berger 2011), which combines the notion of CLV with Social Network Analysis (SNA) techniques. The broad thesis of this stream of work is that developing CLV from a network perspective can lead to additional revenue for the network members, increased customer referrals and improved network relationships among network members. Finding more efficient ways to manage and invest in business network, therefore, is considered as a crucial task for organizations. Little research, however, has studied the networking effect while measuring CLV.

In this research we present a network-based customer lifetime value model, developed in cooperation with a leading British retail bank, which is then applied on transactional data comprising some 900 million business transactions. Existing models are idiosyncratic in nature, driven in good part by the data available to them (Hosseini and Albadvi 2010; Klier et al. 2014; Weinberg and Berger 2011). Here we take the spirit of Klier et al. (2014), developing a more general model applicable to transaction oriented environments: We also focus on firm-to-firm relationships, which is novel in the literature to date. The analytical objectives of the work are to develop (predictive) methods for: (1) Understanding client value by determining the most highly connected and most influential players within the cluster(s) of organizations who transact with each other; and (2) classifying organizations in a relational manner. We address these objectives by using Social Network Analysis (SNA) techniques to model and calculate the Connected Customer Lifetime Value (CCLV), which we define as the present value of the first-order neighbors' influence on the cash flow generated by the focal customer within a network.

In achieving the above, the paper is structured as follows. Section 2 reviews the works related to network analysis as well as different CLV models. Section 3 discusses the development of our Connected Customer Lifetime Value (CCLV) model. The model was developed using a Design Science Research (DSR) approach (Gregor and Hevner 2013; Peffers et al. 2007) though, for reasons of space, we do not discuss this in detail. Section 4 explains the research setting and the data used in this research, discusses the result and their potential impact on the banking decisions related to valuation of its customers. Section 5 concludes the work and notes future work that we will undertake in advancing the state-of-the-art.

Network Analysis and Customer Lifetime Value in the Literature

This section provides an overview of Social Network Analysis (SNA) and its core techniques. Traditional CLV models are reviewed as a platform to discuss recent advances in the state-of-the-art that model 'connectedness' including the impact of social media and the case of referrals (Weinberg and Berger 2011).

A Brief Primer on Social Network Analysis (SNA)

A social network consists of a set of 'entities' and the 'relation(s)' between those entities (Butts 2008). Social network analysis (SNA) is the field of study that is concerned with mapping and measuring of relationships and the flow of values between the nodes that form the network. Nodes can be people, groups or organizations (etc.), while the edges (connections) represent relationships between the nodes (Butts 2008; Kiss and Bichler 2008). Relationships can also take many forms including transactions, information flows, friendships etc. Social Network Analysis (SNA) techniques can help in determining the influential nodes within a network. By presenting a business network using a graph, SNA techniques have the capability to determine highly connected centralized hubs, hence, influential nodes (Chen et al. 2009).

Of importance here, social network analysis encompasses the study of centrality and topological ranking measures, which are used to identify the most valuable vertices within a network (Kiss and Bichler 2008). Centrality measures that are relevant to identifying the most central and influential customers include:

- *Degree Centrality*: In its simplest definition, this represents the number of edges attached to a node. This measure can be divided into in-degree, representing the number of edges arriving to the node, and out-degree, representing the number of edges initiated from the node.
- *Closeness Centrality*: A node is considered as central if it has the shortest path to all other nodes in a network. Identifying the central nodes can improve communications with a network.

- *Betweenness Centrality*: This measure of centrality quantifies the importance of a node (A) as a bridge between two other nodes (B) and (C) when the shortest path between (B) and (C) goes through (A). This means that, node (A) can control the interaction between nodes (B) and (C).
- *Eigenvector Centrality*: Here, the importance of a node (A) is measured by the centrality of the nodes connected to (A). In other words, this measure of centrality means that links with influential people make you more powerful than links with powerless people. PageRank provides a popular and salient example of an Eigenvector method, commonly used to rank the importance of websites via their link structure and can be used to rank the nodes in any network that represent business or social domains (Gleich 2015).

Community structure is an important feature of complex network, which occurs when nodes cluster into tightly knit groups with a high density of within-group connections and a lower density of between-groups connections (Chen et al. 2009). The importance of community detection stems from the fact that it helps in understanding complex systems by de-structuring complex networks into smaller ones with centralized hubs. Moreover, using community detection algorithms enables the detection of nodes that act as bridge between one community and another. Many community detection algorithms exist (Blondel et al. 2008; Chen et al. 2009; Ronhovde and Nussinov 2009; Wang et al. 2007) and, for the interested reader, a review can be found in Lancichinetti and Fortunato (2009).

Customer Lifetime Value (CLV) without Connectedness

Customer Lifetime Value (CLV) is a financial measure that assesses the present value of the cash flow generated by a customer minus the cost of acquisition or retention of that customer – costs such as discounts or promotions (Kumar 2010; Zhang et al. 2016). CLV is generally considered as a trusted metric to measure customer performance in the Customer Relationship Management (CRM) field (Venkatesan and Kumar 2004). The noted benefits of CLV are: (a) The ability to identify the importance of each customer to the organization; (b) the prediction of whether or not it is profitable to acquire new customers or retain existing ones (Blattberg et al. 2009; Feiz et al. 2016); (c) more effective allocation of resources to customers more and improved information on how to develop long-term customer relationships (Tavakolijou 2012); and (d) the ability to predict the probability of customers to defect to competitors in the future (Ferrentino et al. 2016).

In measuring CLV, the standard approach is to estimate the present value of the net benefit to the firm from the customer – generally taken as the revenues from the customer minus the firm's costs in maintaining and developing the relationship with the customer over time (Borle et al. 2008). Many studies have proposed variations on CLV, but the underlying structure is similar (Hosseini and Albadvi 2010). As a representative example, (Berger and Nasr 1998) proposed the following:

$$\boldsymbol{CLV} = \sum_{i=0}^{n} \pi(t)\frac{1}{(1+d)^i} \tag{1}$$

Where:

$\boldsymbol{\pi(t)}$: is the function of customer profits according to time t

i: is the period of cash flow from customer transaction

n: is the total number of periods of customer transactions

d: is the discount rate

$$\boldsymbol{CLV} = \left\{GC * \sum_{i=0}^{n}\left[\frac{r^i}{(1+d)^i}\right]\right\} - \left\{M * \sum_{i=1}^{n}\left[\frac{r^{i-1}}{(1+d)^{i-0.5}}\right]\right\} \tag{2}$$

Where:

GC: is the (expected) yearly gross contribution margin per customer. It is, therefore, equal to revenues minus cost of sales

M: is the (relevant) promotion costs per customer per year

n: is the length, in years, of the period over which cash flows are to be projected

r: is the yearly retention rate, i.e., the proportion of customers expected to continue buying the company's goods or services in the subsequent year

d: is the yearly discount rate (appropriate for marketing investments)

Connected Customer Lifetime Value (CCLV) with Connectedness

Though CLV is of immense value and widespread in use, the concept does not consider the effects (and potential value) of networking among firms (Klier et al. 2014). This gap has motivated research to investigate the significance of considering the customer's surrounding network. Neighbors in a network of customers can refer products to each other. Also, social influence can help companies acquire new customers at relatively low acquisition costs (Klier et al. 2014), and more profitable customers in terms of long-term relationship (Schmitt et al. 2011; Villanueva et al. 2008). Additionally, purchase decision and loyalty can be highly affected by social influence (Nitzan and Libai 2011; Weinberg and Berger 2011). Consequently, discovering the influencers (customers with high connectivity) in a company's network is considered as a crucial task prior to any marketing-related decision making.

One recent model that considers the connectivity of among customers is proposed by Weinberg and Berger (2011), which identifies two kinds of social influence – Customer Referral Value (CRV) and Customer Social Media Value (CSMV). CRV is an important aspect of social influence that affects CLV, as it is the word-of-mouth referral that can lead new customers to buy a product/take-up a service. CSMV is another factor that can cause non-direct cash flow in the network through social media engagement depicted in a form of Twitter tweets, Facebook posts or communities discussion; thus, can affect and change the value of CLV (Weinberg and Berger, 2011). Considering these two aspects, Connected Customer Lifetime Value can be calculated as:

$$CCLV = CLV + CRV + CSMV \tag{3}$$

Where:

$$\boldsymbol{CSMV_i} = CLV_i * \left([1 + SM_{i1}] * [1 + SM_{i2}] * \dots * \left[1 + SM_{ij}\right] * \dots * \left[1 + SM_{iJ}\right] - 1\right)$$

Where:

$\boldsymbol{SM_{ij}}$ is the impact of social media j (Twitter, Facebook, Forums, Communities and Blogs) on customer i

Another approach for calculating CLV while considering the value of network is presented by Hosseini and Albadvi (2010). Here, the value network is divided into tangible (goods and services etc.) and intangible exchanges (knowledge and information that supports the take-up of goods/services). In this research, Network Customer Lifetime Value (NCLV) is defined as:

$$\boldsymbol{NCLV_i} = CLV_i + \sum_{j=1}^{n} \alpha_{ij}\, NRV_{ij} \qquad i \neq j, \qquad j = 1, \dots, n, \qquad 0 \leq \alpha_{ij} \leq 1 \tag{4}$$

Where:

$\boldsymbol{NCLV_i}$ is the network customer lifetime value of customer i

$\boldsymbol{CLV_i}$ is the customer lifetime value of customer i

$\boldsymbol{NRV_{ij}}$ is the network relationship between value between customers I and j

$\boldsymbol{\alpha_{ij}}$ is the importance of NRV_{ij} from focal company's point of view

$\boldsymbol{n}$ is the total number of customer

Last, a study by (Klier et al. 2014) has introduced a different method to calculate CLV in accounting for mutual network effects among the members of a network. The approach may be summarized thus:

$$\boldsymbol{CLNV} = Present\ value\ of\ individual\ cash\ flow + present\ value\ of\ \Delta\, network\ contribution$$

Where:

$\Delta\,\boldsymbol{network\ contribution}$ can be positive or negative depending on the customer contribution to the network. A customer can have positive $\Delta\, network\ contribution$ when the cash flow induced to the network by other members is depending on the cash flow generated by this customer. Conversely, when a cash flow of certain customer is dependent on other members, then $\Delta\, network\ contribution$ for this customer is negative.

The calculation of $\Delta\, network\ contribution$ is dependent on certain variables being computed first. For example, it is necessary to compute the probability that the influence (cash flow) exerted by customer i on customer j will actually lead customer j to make a purchase. Additionally, a discount rate (d) and a weighing factor α are important for the calculation of the final value of CLNV.

Modeling a Connected Customer Lifetime Value (CCLV)

In picking up on the state-of-the-art above, the work here is set in the context of a research project developing network-based analytic and probability-of-default models, alongside related information products in the financial services sector. The work is predicated on the observation that funding decisions made by financial service institutions have unseen network effects – that is, decisions related to funding Firm X have potential knock-on effects for firms related to Firm X. The specific work here was carried out in collaboration with a major UK bank on an anonymized dataset, initially comprising some 900 million inter-firm transactions mediated by the bank (which have been heavily anonymized).

Traditional customer lifetime value models can be seen from the network perspective as the study of the relations between a firm and its clients – as illustrated in Figure 1.a. In such a case, the source node will be one of the clients and the destination node will always be the firm valuing its clients (e.g., ego networks). Here, we are interested in inter-firm relationships, where the form of the relationship (a transaction) is mediated by a third party (a bank) – the bank essentially facilitating the financial manifestation of the inter-firm relationship as illustrated at Figure 1.b. For the study, the mediation of the third-party is immaterial, aside from the fact that all the firms in the dataset are clients of that third-party.

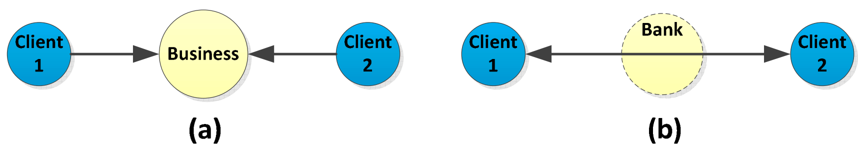

Figure 1. Different types of transactions among firms

As we have noted earlier, the variables used in prior studies vary in accordance with what is available to researchers and what is suitable for the focus of study. In addition, given that we are looking at inter-firm relations that are mediated, variables that are traditionally used in CLV calculations are neither transparent nor available – e.g., costs for promotion, customer acquisition and retention and referral and/or social media factors. Consequently, we have taken the spirit of emerging models (building on the work of (Klier et al. 2014) in particular) and developed one particularized to the domain of problems that we are dealing with.

To explain the model we present it first in the context of a simple fictitious but representative example. Table 1 lists the sample transactional data (the source, the destination and the amount of money transacted (cash flow)) among 8 customers: Figure 2 shows the network representation of the data. The bank (which can be seen as 'a view from nowhere) values its customers by quantifying the cash flow generated by the customer while considering the influence of the customers on each other.

From customer	To customer	Amount
1	2	50
1	5	30
1	6	50
2	4	10
2	3	40
2	1	40
3	7	20
3	2	20
3	8	50
4	2	30
5	1	40
6	1	20
7	3	60
8	3	30

Table 1. Sample transactional data

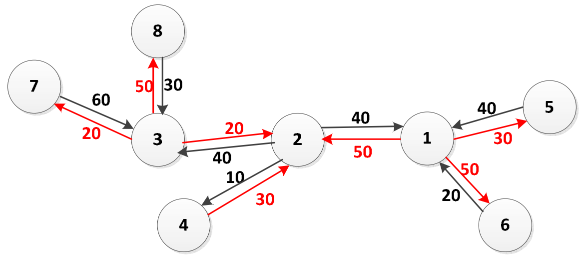

Figure 2. Network representation of the transactional data

In this network, the average weighted network influence of customer (i) on customer (j) compared to other influencers on customer (j) is calculated by dividing the individual influence of the customer (i) by the total average influence on customer (j). Table 2 presents an example of how to calculate the average weighted network influence between customers (1) and (2).

From customer (1) on customer (2)	From customer (1) on customer (2)
50+20+30 = 100/3 = 33.33 50/33.33 = **1.5**	40+40+20 = 100/3 = 33.33 40/33.33 = **1.2**

Table 2. Average weighted network influence

Table 3 shows the matrix representation of influence among all the customers in the sample data. In addition, the table also displays the sum of the columns, which is used to standardize the values in the matrix as displayed in Table 4. Table 4 itself shows how the most influential node (customer) is calculated by summing the rows and dividing the individual results by the total of the sum of the rows. The values in the rightmost column of Table 4 represent the Connected Customer Lifetime Value (CCLV).

From \ To	1	2	3	4	5	6	7	8
1		1.5			1	1		
2	1.2		0.92	1				
3		0.66					1	1
4		0.66						
5	1.2							
6	0.6							
7			1.38					
8			0.69					
Sum of the columns	**3**	**2.82**	**2.99**	**1**	**1**	**1**	**1**	**1**

Table 3. Matrix representation of the influence

From \ To	1	2	3	4	5	6	7	8	Sum of Rows	Standardized sum of row (Most influential Node) (CCLV)
1		0.532			1	1			2.53	2.53/ 8 = **0.32**
2	0.4		0.308	1					1.71	1.71/ 8 = **0.21**
3		0.234					1	1	2.23	2.23/ 8 = **0.28**
4		0.234							0.23	0.23/ 8 = **0.03**
5	0.4								0.40	0.40/ 8 = **0.05**
6	0.2								0.20	0.20/ 8 = **0.03**
7			0.462						0.46	0.46/ 8 = **0.06**
8			0.231						0.23	0.23/ 8 = **0.03**
Total of the sum of the rows									**8.00**	

Table 4. Standardized matrix of influence and calculating most influential node by calculating

standardized sum of row

The values in this matrix show that the influence is equal to 1 when there are no other influencers on the customer. Also, the rightmost column (CCLV) shows that customer (1) is the most influential in the network, then customer (3) and then node customer (2).

Datasets, Results and Discussion

Datasets

In moving from the conceptual to the empirical, we apply the above model to the transactional data provided by our research partner. This data covers business-to-business transactions between Small-to-Medium Enterprises (SMEs) and also includes the Bank's standard value classification of its customers (noted as ValCat in the tables below). The Bank's value classification is categorical ranging from D to A+. From a technical perspective, the computational work was undertaken on a Spark server with an 10 node Hadoop cluster hanging off that. The analysis was performed using the Spark SQL and GraphX libraries. The first step was to only include the transactions where only the customers of the bank are involved in the transaction. Second, it was necessary to exclude the transactions where the customers are moving money between their own accounts (self-paying). Third, we removed the firms such as utility and telecommunication firms as they represent 'false hubs', since every business has to transact with them and they skew outcomes. Approximately 300 million inter-firm business transactions remained. The resulting transactional data is modeled as a scale-free network where the nodes (vertices) represent the organizations and the directed weighted edges represent transactions among those organizations. Scale-free networks usually contain 'hubs' with high degree of centrality (Kiss and Bichler 2008).

Most Influential Nodes (Top 20)

Table 5 presents the Top 20 influential customers in the transactional data network based on the CCLV values. In addition, it provides a comparison between the CCLV and Bank's own valuation of its customer. For completeness, centrality measures are also calculated, including those originating from the customer (out-degree), coming to the customer (in-degree) and the sum of these numbers (degree centrality).

	Node (Anonymised Customer ID)	CCLV (Most Influential Node)	Business Category & ValCat		Degree Centrality			No of Trans (In and Out)	Sum of Amount Cash Flow (In and Out)
			Category	ValCat	Out-degree	In-degree	Degree		
01	9737233578753977	0.0023657041	MM	A+	1,520	1,502	3,022	44,299	£363,778,533
02	9737233577051063	0.0020277463	Missing*		1,477	14	1,491	19,803	£2,279,683,601
03	1964933484975297	0.001934808	MM	A+	947	329	1,276	21,523	£33,154,237
04	9737233564601535	0.001926359	MM	A+	803	657	1,460	27,256	£953,122,361
05	7760150167739034	0.0018672164	SME	A+	554	371	925	25,426	£7,166,569
06	1184727448856452	0.0018165228	SME	A+	1,870	1,913	3,783	119,324	£36,543,368
07	3316428431849140	0.0015546055	MM	C	950	146	1,096	3,150	£4,727,054
08	9737233574695254	0.0015039119	MM	A+	1,174	461	1,635	16,972	£126,336,918
09	3044772329461175	0.0013180351	MM	D	1,620	26,732	28,352	361,961	£214,329,592
10	9737233532427343	0.0011997499	MM	A+	1,087	10	1,097	9,578	£107,418,271
11	9737233592737223	0.00103922	MM	A+	1128	300	1,428	26,953	£118,388,635
12	6008353554890705	9.885263E-4	MM	C	202	18	220	11,973	£5,380,468
13	9737233555064670	9.800774E-4	MM	A+	633	434	1,067	15,134	£110,365,060
14	9737233530665152	8.617922E-4	MM	A+	1,041	723	1,764	17,037	£157,170,001
15	9556487459311823	8.448943E-4	MM	A+	427	279	706	4,596	£13,918,953
16	5183006448623432	7.942006E-4	MM	A+	271	61	332	9,249	£36,083,006
17	9737233562316476	7.857517E-4	MM	B	635	118	753	11,768	£3,090,350,308
18	2715714432040724	7.773028E-4	MM	A+	312	225	537	18,883	£116,269,852
19	7299138661748686	7.688538E-4	MM	D	286	3	289	2,482	£70,690,695
20	3323635188960776	7.266091E-4	Missing*		583	57	640	9,399	£204,722,136

* Missing observations from the ValCat data provided by the bank

Table 5. CCLV for most influential nodes compared to ValCat and Degree centrality

From a descriptive point of view, the points of interest are as follows. First, there is a visible correlation between the CCLV ranking and ValCat value – most of the top 20 nodes are valued by the bank as A+ customers. Second, discrepancies exist, however, and the firm ranked 9th provides a striking one. This firm is very highly connected with out-degree of 1,620, an in-degree of 26,732 and a centrality score of 28,352 – the firm also has a relatively high cash flow of £65,335,780. Third, the firm ranked 17th has a significant cash flow of more than 3 billion GBP but is valued as B. Degree centrality measures are modest. Last, it can be concluded here that the ValCat is crude and does not accurately rank the customers according to their importance to their surrounding network. Thus, CCLV provides more accurate measure to rank the customers in a network.

Least Influential Nodes (Bottom 20)

Tabe 6 presents the bottom 20 influential customers in the bank's transactional data network based on the CCLV values. For consistency, the comparison between the CCLV and ValCat is shown, as are the centrality measures.

	Node (Anonymised Customer ID)	CCLV (Least Influential Node)	Business Category & ValCat		Degree Centrality			No of Trans (In and Out)	Sum of Amount Cash Flow (In and Out)
			Category	valcat	Out-degree	In-degree	Degree		
01	7760150148859210	0.0	SME	C	5	1	6	38	£29,023
02	0888254039226130	0.0	Missing*		Did not calculate**			0	£0
03	0690272583058668	0.0	Missing*		Did not calculate**			0	£0
04	7913952738710923	0.0	SME	B	25	12	37	190	£1,086,750
05	8309135282123031	0.0	SME	A+	16	12	28	96	£46,750
06	5584821899727824	0.0	SME	A+	6	60	66	189	£274,065
07	3683757165980692	0.0	SME	C	9	12	21	201	£321,724
08	8241990545812517	0.0	SME	B	4	1	5	54	£642135
09	6746568044338914	0.0	SME	B	9	11	20	98	£23,280
10	2473614834465458	0.0	Missing*		Did not calculate**			4	£36,320
11	1739074866884483	0.0	SME	A+	3	6	9	168	£438,489
12	1960646785381956	0.0	SME	C	Did not calculate**			255	£423,721
13	9386193073253650	0.0	SME	A+	22	16	38	336	£395,101
14	8402089455937756	0.0	SME	C	5	8	13	238	£176,645
15	1741462932647747	0.0	SME	C	1	1	2	21	£283,400
16	7896916623965298	0.0	SME	C	14	6	20	30	£5,856
17	4035855558378531	0.0	SME	C	12	2	14	23	£27,792
18	6196310905978599	0.0	SME	A+	17	35	52	350	£951,235
19	3190225371870381	0.0	SME	A+	14	37	51	497	£1,775,666
20	2954239842028123	0.0	SME	C	7	32	39	140	£140,722

* Missing observations from the ValCat data provided by the bank
** Degree centrality did not calculate on GraphX in Spark for those nodes

Table 6. CCLV for least influential nodes compared to ValCat and Degree centrality

From a descriptive point of view, the points of interest here are as follows. First, there is much less of a visible correlation between the CCLV score (which is zero for all firms) and the Bank's valuation. Second, all firms here are Small-to-Medium Enterprises (SMEs) and, unsurprisingly, they are much less connected. Last, discrepancies again exist, however, and the firm ranked 11th provides an illustrative example. This firm is valued as A+ even though its out-degree is 3, in-degree is 6 and centrality is only 9 (very low connectivity). In addition, the total number of transactions for this firm is only 168 and the cash flow is very low across the 3 years span of the data.

Discussion

The CCLV model adds a potentially valuable tool to the analytic arsenal. From the discrepancies between the CCLV scores and the Bank's traditional valuation, it is fair to conclude that the outcomes tell different stories – both of which have value in the domain. The CCLV allows a ranking of importance, based on the influence that a firm has amongst its first-order relations. Most importantly, modeling in network form makes clear that decisions made re one firm has tangible knock-on effects for others. In addition, it opens up opportunity for new information products and marketing strategies that, for example:

1. Exploit and strengthen synergies between firms.
2. Enable the better allocation of resources, depending on a customer's individual network value.
3. Exploit a different form of customer segmentation based on relations.

The review of the literature shows variance in the way that emerging CCLV models are constructed – that variance stems, in good part, from data that is available to support the construction of the model. Having developed the model in conjunction with our partner, our only claim at this point is that the model is fit for purpose and is able to calculate CCLV with the minimum information available. In the context of fledgling work in the area, it provides contribution in the form of applicability to transaction-oriented environments and a focus on firm-to-firm relationships (which is novel in the literature to date).

Our future research in this area will seek to improve upon the model presented here, based on feedback from use within the context of application. First, though we cannot model the probability of future transactions in the manner of Klier et al. (2014), providing a proxy in the form of the centrality scores is a useful contribution. Second, given the provision of additional firmographic and product holding data by our partner, we will examine how the proposed CCLV model can benefit from benchmarking against common performance indicators. For example, companies of a certain size, sector and geographical location with similar amount of connections could have certain banking products. This allows us to address potential weaknesses in comparison to the ValCat measure, which is not transaction oriented and hides the factors used to derive it. Last, we will time series the data and use machine learning techniques to predict the likelihood that firms will invest in products/services based on their relational standing (in the sense of a contagion model).

Summary and Conclusions

Since its inception, Customer Lifetime Value (CLV) has become increasingly important as a marketing metric – not least because it can positively affect current and future performance. A significant strength of CLV analysis is that it can be used to predict the future profitability of clients, potentially enabling more appropriate marketing strategies and decisions relating to customers. The downside of the approach is that it ego-based and does not reflect the (dynamic) networks that reflect the manner in which firms now operate. Recognizing this, fledgling research has emerged that seeks to add a 'connected' element to the more general CLV approach.

Here, we have sought to add to the 'connected' research base, developing a Connected Customer Lifetime Value model based on an analysis of transactions in the financial services domain. The model was applied to a significant number of transactions between firms using a modern open source computing infrastructure (Spark plus Hadoop). We have illustrated the outcomes of the application of the model via a 'top and bottom 20' listing of firms in relation to their connected value. In positive terms, application of the model allows our research partner to see the network implications of decisions they make with respect to customers and opens up an arena for innovation re network-based products and services. We have sought to add to the research base claim at this point only that the model is fit for purpose and is able to calculate CCLV with the minimum information available. Key areas for improvement include adding probability of future transaction measures, benchmarking against company performance data and applying machine learning techniques to assess potential product/service contagion.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (Grant No. EP/L021250/1)

REFERENCES

Baum, M., and I. Singh. 2008, "Winning the Customer Service Battle," *Chicago: AT Kearney, Inc* (5:.

Berger, P. D., and N. I. Nasr. 1998, "Customer Lifetime Value: Marketing Models and Applications," *Journal of Interactive Marketing* (12:1), pp. 17-30.

Blattberg, R. C., E. C. Malthouse, and S. A. Neslin. 2009, "Customer Lifetime Value: Empirical Generalizations and some Conceptual Questions," *Journal of Interactive Marketing* (23:2), pp. 157-168.

Blondel, V. D., J. Guillaume, R. Lambiotte, and E. Lefebvre. 2008, "Fast Unfolding of Communities in Large Networks," *Journal of Statistical Mechanics: Theory and Experiment* (2008:10), pp. P10008.

Borle, S., S. S. Singh, and D. C. Jain. 2008, "Customer Lifetime Value Measurement," *Management Science* (54:1), pp. 100-112.

Butts, C. T. 2008, "Social Network Analysis: A Methodological Introduction," *Asian Journal of Social Psychology* (11:1), pp. 13-41.

Chang, W., C. Chang, and Q. Li. 2012, "Customer Lifetime Value: A Review," *Social Behavior and Personality: An International Journal* (40:7), pp. 1057-1064.

Chen, D., Y. Fu, and M. Shang. 2009, "A Fast and Efficient Heuristic Algorithm for Detecting Community Structures in Complex Networks," *Physica A: Statistical Mechanics and its Applications* (388:13), pp. 2741-2749.

Feiz, S., A. RamezaniGhotbabadi, and Z. BteKhalifah. 2016, "Customer Lifetime Value in Organizations," *Asian Journal of Research in Social Sciences and Humanities* (6:5), pp. 53-58.

Ferrentino, R., M. Cuomo, and C. Boniello. 2016, "On the Customer Lifetime Value: A Mathematical Perspective," *Computational Management Science* (13:4), pp. 521-539.

Gregor, S., and A. R. Hevner. 2013, "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-356.

Gleich, D. F. 2015, "PageRank Beyond the Web," *SIAM Review* (57:3), pp. 321-363.

Gupta, S., D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, and S. Sriram. 2006, "Modeling Customer Lifetime Value," *Journal of Service Research* (9:2), pp. 139-155.

Hosseini, M., and A. Albadvi. 2010, "Customer Value Network Analysis for Improvement of Customer Life-Time Value Computation," *International Journal of Electronic Commerce Studies* (1:1), pp. 15-24.

Kiss, C., and M. Bichler. 2008, "Identification of influencers—measuring Influence in Customer Networks," *Decision Support Systems* (46:1), pp. 233-253.

Klier, J., M. Klier, F. Probst, and L. Thiel. 2014, "Customer Lifetime Network Value," *Proceedings of the Thirty Fifth International Conference on Information Systems ICIS, Auckland, NZL.*

Kumar, V., K. N. Lemon, and A. Parasuraman. 2006, "Managing Customers for Value: An Overview and Research Agenda," *Journal of Service Research* (9:2), pp. 87-94.

Kumar, V. 2010, "A Customer Lifetime Value-Based Approach to Marketing in the Multichannel, Multimedia Retailing Environment," *Journal of Interactive Marketing* (24:2), pp. 71-85.

Lancichinetti, A., and S. Fortunato. 2009, "Community Detection Algorithms: A Comparative Analysis," *Physical Review E* (80:5), pp. 056117.

Mzoughia, M. B., and M. Limam. 2015, "An Improved Customer Lifetime Value Model Based on Markov Chain," *Applied Stochastic Models in Business and Industry* (31:4), pp. 528-535.

Nitzan, I., and B. Libai. 2011, "Social Effects on Customer Retention," *Journal of Marketing* (75:6), pp. 24-38.

Peffers, K., T. Tuunanen, M. A. Rothenberger, and S. Chatterjee. 2007, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45-77

Ronhovde, P., and Z. Nussinov. 2009, "Multiresolution Community Detection for Megascale Networks by Information-Based Replica Correlations," *Physical Review E* (80:1), pp. 016109.

Schmitt, P., B. Skiera, and C. Van den Bulte. 2011, "Referral Programs and Customer Value," *Journal of Marketing* (75:1), pp. 46-59.

Tavakolijou, M. 2012, "A Model to Determine Customer Lifetime Value in Iranian Banking Industry," .

Venkatesan, R., and V. Kumar. 2004, "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing* (68:4), pp. 106-125.

Villanueva, J., S. Yoo, and D. M. Hanssens. 2008, "The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth," *Journal of Marketing Research* (45:1), pp. 48-59.

Wang, X., G. Chen, and H. Lu. 2007, "A very Fast Algorithm for Detecting Community Structures in Complex Networks," *Physica A: Statistical Mechanics and its Applications* (384:2), pp. 667-674.

Weinberg, B. D., and P. D. Berger. 2011, "Connected Customer Lifetime Value: The Impact of Social Media," *Journal of Direct, Data and Digital Marketing Practice* (12:4), pp. 328-344.

Zhang, H., X. Liang, and S. Wang. 2016, "Customer Value Anticipation, Product Innovativeness, and Customer Lifetime Value: The Moderating Role of Advertising Strategy," *Journal of Business Research* (69:9), pp. 3725-3730.